A Survey on Challenges of Semantics Application in the Internet of Things Domain

Marina Harlamova[1] Mārīte Kirikova[2], Kurt Sandkuhl[3]
[1, 2] *Institute of Applied Computer Systems, Riga Technical University, Latvia*
[3] *Faculty of Computer Science and Electrical Engineering, University of Rostock, Germany*

Abstract – The Internet of Things (IoT), a global Internet-based system of computing devices and machines, is one of the most significant trends in the information technology area. An accepted unified communication approach would be a prerequisite for its mass adoption. Semantic technologies (Semantic Web) have been advocated as enablers of unified communication. However, while there are particular advancements in research on application of Semantic Web in the IoT domain, the dynamic and complex nature of the IoT often requires case specific solutions hard to be applied widely. In the present survey, the semantic technology challenges in the IoT domain are amalgamated to provide background for further studies in the use of semantic technologies in the IoT.

Keywords – Internet of Things, Semantic Web, semantics.

I. INTRODUCTION

At its core, the idea of the Internet of Things (IoT) is straightforward – to connect devices to the Internet to be able to exchange information anytime from anywhere to any device. Ability to obtain autonomous sensory device readings is changing the way how business could work and services could be provided in future. This trend creates a range of unique new services and products in a variety of areas, such as monitoring in health sector, logistics solutions, smart homes, governing of smart cities and other areas. The Internet as we know it today will possibly slowly evolve into a platform where classical computer networks and interconnected objects are joined together, thus enabling new ways of communication and interaction [1]. The Internet will form the backbone of global information exchange and provide opportunities for companies working in the IT and telecommunications sector to leverage the connectivity of both physical devices and virtual structures. Studies by Ericsson predict that the number of objects in the IoT will reach 50 billion by 2020 [2].

An important factor to be considered in the IoT domain is the stakeholders involved in data exchange. In the typical Internet and WWW use experienced during the past decade, in an end-to-end communication sessions, both data providers and receivers are human beings [2]. In the IoT, the main roles are taken by things or objects that are generating the majority of data. For the IoT it means that essential aspects to develop and consider are high level information processing, anomaly detection, and one-way communication to a specified location. Thus, there is a danger of conflicts between IoT requirements to the Internet use and "traditional" Internet applications as traditional Internet use might become inefficient [3]. To enable global machine communication by autonomous information gain, discovery, analysis, and proposal, it is crucial to structure and group the exchanged data properly. Storage, ownership and expiry of data become especially important challenges to handle [4].

Semantic-oriented approaches have already been introduced at the early stages of studying the IoT domain to deal with big data, interoperability and achieve actionable knowledge handling. Semantic Web technology enables devices to "understand" and process data using structured and machine-readable descriptions of resources [5]. Machine-interpretable data descriptions can also show where data originates from, how it can be related to its context, who provides the data, and what the various attributes of the data are [6]. Semantic frameworks help semantically annotate and easily interpret information exchanged among IoT devices. In order to make and maintain effective business decisions, many organisations are leveraging external data sources from Semantic Web, e.g., social networks, media feeds, sensor data, or any other generic published information [7]. Joined knowledge on semantic principles and their application are supposed to improve the standardisation and adaptation processes of IoT devices in the nearest future.

In the paper, a systematic literature review has been conducted to identify commonalities and open issues regarding application of semantic technologies in the IoT domain. In the survey, Semantic Web is often used as a synonym of the term "semantic technology", unless other semantic technologies are considered.

The remainder of the paper is structured as follows. In Section II, the research method is presented. In Section III, the basic concepts regarding application of semantic technologies in the IoT domain are described. In Section IV, the open issues and challenges found in related works are discussed. Section V consists of brief conclusion.

II. SURVEY METHOD

A systematic review helps highlight identified research challenges and guide scientific work into the direction of unresolved problems. The goal of the survey was to review existing research to assess the challenges of application of semantic technologies in the IoT area. To achieve this objective, the following general research questions were stated:

Applied Computer Systems

_____ *2017/21*

- How can we assess conclusions in the available research on implementing semantic technologies to the IoT and summarise subsequent challenges?
- What are the problems in the IoT domain that can be possibly answered by a semantic technology?
- What are the problems derived from the application of each semantic technology?

The list of eligibility criteria was established to allow for a transparent selection process of papers, i.e., to decide if a certain article was to be included or excluded in the survey. The following mandatory criteria for including the article in the survey were defined:

- The article was published in English in a peer reviewed source;
- The article contained comprehensive conclusions on performed reviews or developments;
- At least one of the following additional criteria should be met to include the article to research scope:
  - The study considered the subject of IoT challenges;
  - The study considered the subject of semantic technology challenges (as these challenges could also arise in the IoT domain);
  - The study considered both IoT challenges and semantic technology challenges.

The following criteria were used for excluding the articles from the survey:

- The article did not give an answer to any of the stated research questions (even not partly);
- It was not possible to derive any information on potential future challenges from the article.

Based on questions that had to be answered during the survey and considering the limitations proposed by the exclusion criteria, the search concepts and notions were defined. These concepts were used as look-up criteria in the literature search as follows:

- ("Semantic Web" OR "semantic technology" OR "semantics") AND ("challenges" OR "problems" OR "assessment");
- ("IoT" OR "Internet of Things" OR "Internet-of-Things" OR "machine learning") AND ("challenges" OR "problems" OR "assessment");
- "Semantic Web in IoT" OR "semantic technology in IoT" OR "Semantics in IoT";
- Other possible combinations or derivations of terms described above.

Systematic review guidelines suggest that searching in the title does not always result in all relevant publications. Therefore, the abstract or full-text of articles should also be considered. Thus, the search look-up criteria were applied to the title and abstract of the studies. The search was started with the articles available in SCOPUS, one of the well-recognised indexes of scientific papers. In SCOPUS search, the filter selecting only the document type of "Article" was used. Displayed results were ordered by citation number to ensure inclusion of papers generally acclaimed by the scientific public. Afterwards the same keywords and filters were also applied in ACM Digital Library, EBSCOHOST, IEEE Xplore Digital Library, SpringerLink, and ScienceDirect (see Fig. 1).
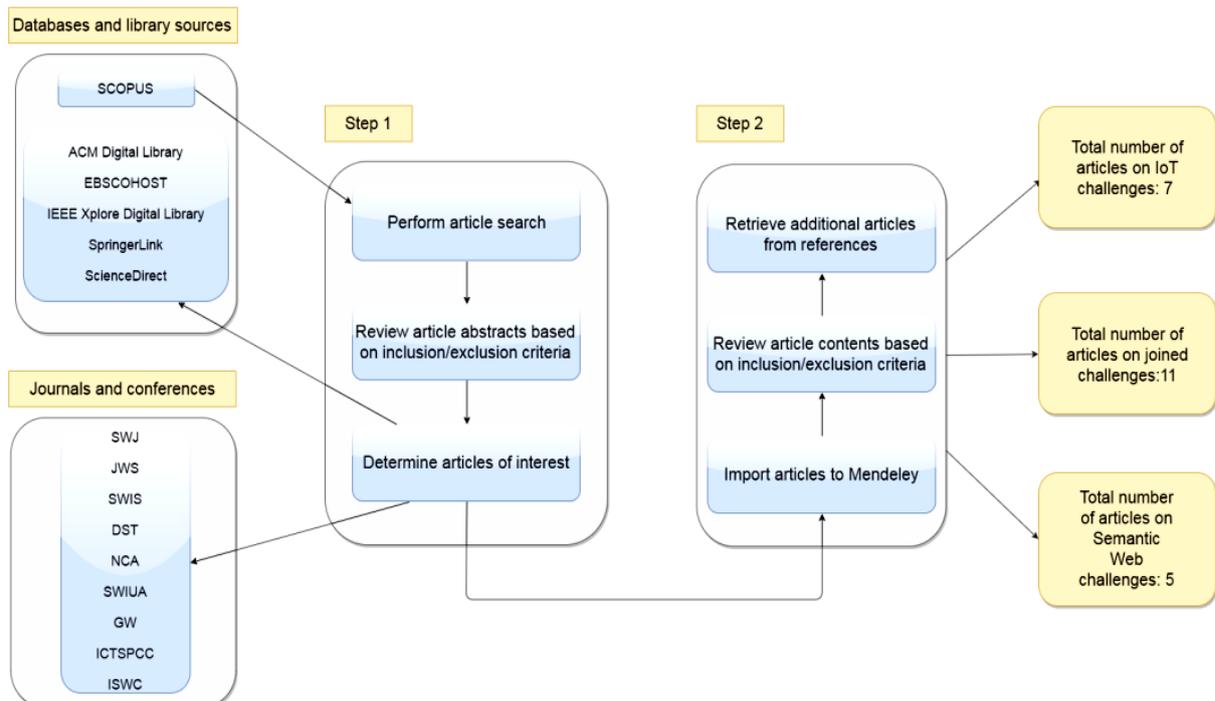

Fig. 1. Approach to retrieving articles during systematic literature search.

Applied Computer Systems

_____ *2017/21*

The following journals were considered: Semantic Web Journal (SWJ); Journal of Web Semantics (JWS); International Journal on Semantic Web and Information Systems (SWIS); International Journal of Distributed Systems and Technologies (DST); Journal of Network and Computer Applications (NCA); and Semantic Web – Interoperability, Usability, Applicability (SWIUA). The proceedings of the following conferences and workshops were considered: 2014 IEEE Globecom Workshops (GW); 11th IEEE Int. Conference on Trust, Security and Privacy in Computing and Communications (ICTSPCC); and International Semantic Web Conference (ISWC). Bibliographical data in the number of 38 preliminary studies were collected and stored in *Mendeley* software platform for further evaluation. An overview of described search approach is shown in Fig. 1.

The systematic literature search and the application of eligibility criteria resulted in the list of relevant articles. A total of 23 papers published from 2002 to 2016 were collected. Of these papers, 7 articles are focused on challenges in IoT area, 5 items are describing challenges in Semantic Web, and 11 articles mention challenges of applying semantic technologies in the IoT domain. Majority of the used papers (21 out of 23) were published in the period 2010–2016 ensuring that their content corresponded to the state of the art in the area and was relevant for this survey.

There were two main groups of research works identified under the topic of Semantic Web for IoT, namely, (1) technical/practical research papers, e.g., development of applicable frameworks in the IoT based on semantic/machine-learning technologies, in some cases including also case studies and practical examples and (2) theoretical/review papers, e.g., studies of recent applications of Semantic Web to the IoT and analysis of ongoing areas of research. For the first group, the proposed semantic models and frameworks that might be applied to real-life objects and observations were analysed. The second group, focusing on existing research, also contained survey papers that presented findings of different approaches and issues in semantics within the IoT. Since the area of semantics in the IoT and IoT as a research field, in general, are at an early stage of progression, the survey scope contained not only papers directly describing semantics in the IoT but also articles on semantic issues with no direct relation to IoT and articles on IoT service challenges and emerging issues with no semantic research aspect. These articles were considered because they might show potential challenges in the use of semantic technologies in the IoT. Thus, three groups of challenges were identified in the papers presented in Table II:

- Challenges derived from existing surveys in the area of IoT with semantics;
- Challenges derived from developed applications and experiments with semantics in the area of IoT;
- Challenges delineated purely from the field of semantics or field of IoT, tendering potential further correlation.

All three groups are important to derive the challenging topics in the application of semantic technologies in the IoT domain.

Analysing and delineating problem areas of the second column of Table I, papers show what problems have emerged during the application of semantic data. Combining the issues derived from papers of the second column of Table I with the survey findings and literature reviews (the first column of Table I) as well as emerging issues in particular fields of research from the papers of the third column of Table I, a strong backbone for the survey paper can be formed, consisting of practical existing challenges and theoretical potential areas of research and improvement. Altogether, from different columns of Table I, 7 papers led to the IoT challenges, 5 papers led to Semantic Web challenges, and 11 papers reported on semantic technology challenges in the IoT domain.

TABLE I
SURVEY PAPER COMPARISON BY THE CONTEXT

Semantic IoT surveys	Applications & experiments	IoT or semantics
[6], [8]	[5], [9], [10], [11], [12], [13], [14], [15]	[1], [2], [3], [4], [16], [17], [18], [19], [20], [21], [22], [23], [24]

III. BACKGROUND

In this section, we describe the basic concepts that will be further used when reporting on different groups of challenges found in the selected papers. We describe the concepts behind the semantic technologies, semantic language and reasoning, current state of IoT, and data exchange in the IoT.

A. Concepts behind Semantics Technologies

Semantic Web is augmenting the human-understandable part of the Web with structured knowledge that is available for machines to process. Usually services of today's Web interface require human intelligence for perception of stored data. However, semantic technologies require large data storage units that machines can independently access, process and act accordingly [16]. It is predicted that successful applications of semantics on computers will create new technologies that use human-readable and structured machine data to assist both humans and machines, also resulting in development of new, previously non-existing services [17]. Increasing machine autonomy and strong demand for technology in daily life explain why many areas of computer science have been involved into development of Semantic Web over the years. Computer science more and more focuses on artificial intelligence. Databases need to be optimised for structured data storage, ensuring transactions and enabling query communication. System sciences provide mobility and reliability of semantic technologies. Existing libraries and documents are reengineered, transformed and indexed to propose a format that can be easily processed by computers.

Semantic communication includes converting normal one-dimensional "raw" data into rich "polyhedral" information units [16]. For instance, a measured value of 23 can be received at a communication end-point as a unit with specified data type "water temperature", unit of measure "°C", exact time "12:45:32", etc. This transformation demands high level

machine processing in real time and effective data storage methods. Computers have to be able to access needed resources at any time. Consequently, the resources have to be open and similarly configured to ensure fault tolerance. Networks that use semantic technologies should be available for on-demand modification. Networks should also support adding or removing devices without communication downtime. All mentioned factors show that a vital component for Semantic Web discipline is ensuring interoperability [17]. Two approaches, as defined by [9], can be used to reach interoperability: full standardisation and partial standardisation combined with artificial intelligence solutions. Standardisation of all developed and developing technologies seems improbable as it requires worldwide agreements between manufacturers and research groups. Partial standardisation combined with a particular level of intelligence provided by semantic recommendations (required formats/languages) is more plausible.

B. Semantic Language and Reasoning

The Semantic Web concept is largely based on the Resource Description Framework (RDF), which uses Internationalised Resource Identifiers (IRI) to enable linking and merging of relations between units from multiple resources in the web. This method was first proposed by Berners-Lee in [25] and included in Linked Data concept. RDF is a platform-independent development that offers different level information representation in a consistent manner. However, by default, information offered by RDF is generic. More advanced languages have been developed on top of RDF to add service or application specific features to base RDF. Once the information is structured in RDF, a query processor engine can be used to find specific knowledge or information. Search can be executed in a simple way similar to keyword search in regular web, or can be performed in an advanced way by defining document content in data trees within documents.

Apart from representation formats, vocabularies and ontologies are vital for structured information transfer in Semantic Web. Vocabularies define relationships and terms used to describe and represent an area of interest. Ontologies, in addition to dictionaries, allow for more expressiveness of the concepts [18]. Ontologies are suitable for sharing, organising and representing knowledge in a formal way [10]. RDF Schema, OWL and similar semantic tools offer vocabulary for representing and designating RDF data. Ideally, a universal ontology could be used to establish variable formal vocabulary to share between communicating entities, especially in the artificial intelligence area of discourse [19]. However, in practice, there is a movement towards a definite set of ontologies that are agreed upon in particular domains instead of only a universal ontology.

Semantic reasoning is another paramount concept in Semantic Web. Reasoning is deduction of facts explicitly not described in a vocabulary or knowledge base using rules or definitions [18]. By means of reasoning, actual machine knowledge can be achieved. Reasoning and inferences have a key role in the development and integration of semantic

ontologies [19]. A reasoning engine (also called a reasoner) is a software tool that executes reasoning based on rules. Rules, in turn, are based on first-order predicate logic or description logic (DL) to derive conclusions from a sequence of statements (premises) [10].

C. Current State of IoT

During a presentation for Procter & Gamble in 1999, a term IoT was proposed by Kevin Ashton. The idea behind IoT was to enable computers to send information about the current state of physical world objects to the web [11]. Object identification was implemented using radio-frequency identification (RFID). Today the term IoT is used to refer to global network of interconnected devices or "things" by means of vast Internet technologies [1]. RFIDs, sensors and actuators, machine-to-machine devices are leveraging into increasing potential of IoT networks for new service and business opportunities. The conceptual flow in IoT is similar to a human-usable Internet concept. In IoT, the main data producers and consumers are things or devices [20]. IoT components must possess the following abilities: (1) be identifiable, (2) be able to communicate, and (3) be able to interact [1].

In most of current implementations, devices that combine aforementioned requirements generally can only operate within boundaries of a network of the same producer and do not possess sharing capabilities between different vendors and their platforms [10]. Knowledge can be utilised only for a particular domain in a particular case of appliance. Full potential of IoT could be achieved when devices are taught to communicate and collaborate with all accessible devices, regardless of vendor or producer.

D. Data Exchange in the IoT

Before discussing ubiquitous data exchange in IoT, the five-layer technical architecture of IoT has to be introduced (see Fig. 2). The bottom layer or perception layer (also called edge layer) can be identified as a hardware or physical layer. This is where sensors are used to collect data. The next, access gateway layer, together with the network layer is used to provide and manage communication between objects and systems in the IoT environment. Further, a middleware layer provides a dynamic interface between hardware and applications. The top layer is an application layer that allows services or applications to integrate or analyse the received information from lower layers. This is an enhanced architecture of current IoT paradigm that was initially built as only a three-level architecture, with the first, third, and fifth layers present [20].

The main goal behind IoT is to create context awareness by enabling the devices, applications and services to respond dynamically to changes or requests from their environments [21]. In flexible IoT systems, data are delivered between message brokers. Brokers are server-side software components that exchange messages between a message transmitter and a receiver. Exchange solutions may contain built-in message routing based on content, as well as message summarisation and extraction [10]. Describing IoT data taxonomy, Quin et al.

Applied Computer Systems

_____ *2017/21*

in [2] classify characteristics of IoT data and propose the following three categories: data generation, data quality, and data interoperability. Data generation considers the aspects of different rate of data generation in the IoT, foreseeable increase of data scalability in the IoT with more and more devices connecting to networks, mobility and dynamicity of IoT data, and heterogeneity of data resulting in varying formats using different vocabularies. Data quality concerns uncertainty as the received information might be missing certain attributes; redundancy as multiple sensors repeat the readings at the same locations; ambiguity as data interpretation can differ for various data consumers; and inconsistency factors such as inaccuracy and data instability over time in the changing environment. Data interoperability contains issues of incompleteness and proposes semantic aspects of data in the IoT as a method to progress towards interoperability.
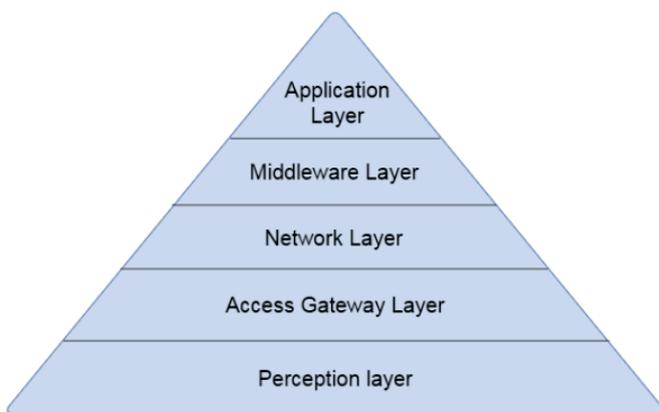

Fig. 2. Five-layer architecture of the IoT.

The idea of Information-Centric Networking (ICN) has emerged as a possible evolution path for future Internet and IoT. This method assumes that data resources are named based on content of the information rather than IP address as in standard networking protocols. Thus, it offers good integration possibilities with developing IoT applications [10]. It is predicted that cloud computing, social networks, and big data research areas may have a strong impact and lead the way of future IoT developments [20].

IV. SEMANTICS IN THE IOT: MAIN CHALLENGES

Challenges of IoT are related to the possibility to consider relevant issues of the physical world, ensuring technical interoperability from technologies to deliver information, and ensuring a possibility for the information to be understood and processed. In this section, the challenges identified from related work which are related to semantic technologies in the IoT are organised in the following six groups: scalability and flexibility; standardisation and reusability; high level processing; data quality; data confidentiality and privacy; and interpretation and synthesis. Table III presents the groups of challenges and the studies that at least partly address them. The subsections of this section offer a brief review of each group of challenges and identify solutions (expected and provided) to address these challenges.

TABLE II
DISCOVERED GROUPS OF CHALLENGES

Source	Scalability and flexibility	Standardisation and reusability	High level processing	Data quality	Data confidentiality and privacy	Interpretation and synthesis
Gubbi et al., 2013 [4]	✓	✓	✓		✓	
Gyrard et al., 2014 [21]		✓	✓			✓
Miorandi et al., 2012 [1]	✓	✓	✓		✓	
Qin et al., 2015 [24]			✓			
Qin et al., 2016 [2]	✓	✓	✓	✓	✓	✓
Tsai et al., 2014 [20]	✓	✓			✓	✓
Weber, 2010 [22]	✓	✓			✓	
Euzenat, 2002 [17]			✓	✓	✓	
Fürber, 2016 [16]		✓		✓		
Martinez-Rodriguez, 2016 [18]	✓	✓	✓	✓		✓
Noy, 2004 [19]		✓				
Zaveri et al., 2014 [23]	✓	✓		✓		
Barnaghi et al., 2012 [6]	✓	✓	✓	✓	✓	✓
Christophe, 2012 [13]	✓		✓	✓		
Darbari et al., 2015 [15]		✓				
Kiljander, 2014 [11]	✓	✓	✓			
Kotis et al., 2013 [9]		✓				
Maarala et al., 2016 [10]	✓	✓		✓		
Perera et al., 2014 [8]	✓	✓	✓	✓	✓	✓
Perera et al., 2012 [3]	✓		✓			
Pfisterer et al., 2011 [5]	✓	✓				
Song et al., 2010 [14]		✓				
Wang et al., 2012 [12]	✓	✓		✓		

✓ - Issue is mentioned and analysed

✓ - Issue is mentioned, addressed and solution proposed

A. Scalability and Flexibility

Properties of highly dynamic and distributed systems consisting of a large number of communicating objects indicate that scalability is to become one of major research problems and requires development of flexible and extendable solutions. The research in this area has to consider the following issues: automated (or semi-automated) annotation of available resources, semantic association discovery and analysis (resource connection or activation), and efficient solutions to create, analyse and explore linked IoT data of various resources [6]. Some studies in the IoT domain address the rising problem of ensuring IoT scalability [21]. Using known P2P (peer-to-peer) methods for IoT systems promises good scalability possibilities [22]. For applications utilising semantic notations, some data engineering has been done to

meet the scalability requirements in IoT. Pfisterer et al. in [5] proposes a SPITFIRE prototype that enables scalability by avoiding registry of semantic entities. Research conducted by Wang et al. [12] describes the idea of providing IoT information through standard service interfaces that resonates with service-oriented computing paradigm, and provides scalability respectively. The idea of "sensing as a service" represents a scalable way to access the sensor data through standard services. Recent work by De et al. focuses on a semantic description model for services exposed by the IoT resources [26]. Scalable search by means of distributed framework of nodes is proposed in [13], hence ensuring satisfactory management mechanisms for IoT data networks.

B. Standardisation and Reusability

Standardisation is a vital area of concern in any data engineering field as it infers stable quality levels over time and steadiness against unexpected faults. Technical standards are based on consensus of different parties involved in development or production of a given service or device. It is essential to define and follow a standard specification guide so that further enhancements can be added to existing solutions without significant effort [8]. The idea of standardisation of IoT interconnected devices on a global scale is improbable at this point; however, certain regulatory approaches are aiming at standardisation [22]. The European Commission provides open framework guidelines for development of IoT devices for ensuring that such principles as verticality, ubiquity, and technicity are well aligned. However, they only apply to Europe and by no means can be considered a global measure. Tsai et al. in [20] mentions global standardisation as an important issue for future IoT development. A possibility to avoid standardisation problems is the use of extending standards with adapters to other standards. However, certain drawbacks take place. Firstly, one standard has no direct control over other standards, which means that changes for one standard will not automatically be propagated to other standards. Secondly, in order to support interoperability among several standards, a large number of adapters have to be developed, which is clearly inefficient [14]. There are distinct missing standardisation activities related to data models, ontologies, and data formats to be used in IoT applications for service-level interfaces and protocols [1]. Machine-to-Machine Measurement (M3) framework is offered to supplement existing semantic standards by adding common format, nomenclature and methods for data interpretation [21]. Semantic approach is aimed at resolving the issue of lack of standardisation by introducing common ontologies, data models, and vocabularies; however, currently the application methods are non-unified, complicated, and require further improvement. Evolution of Semantic Web has produced a large number of semantic models and frameworks that do not always possess a means to be easily applied and tailored for the IoT paradigm. Such a situation is a great playground for experienced users, but for beginners it could lead to misunderstanding of the benefits of structured data representation [18]. Developing comprehensive technical architecture supports the goal of defining a common data handling structure in the IoT. Kiljander et al. [11] propose and evaluate particular semantic interoperability architecture.

The main contribution of RDF and OWL semantic notations and ontology aside from standardisation is that they also support information reusability [11]. Ontologies contain several generic key components such as relations, classes, attributes, individuals, function terms, restrictions, rules, axioms, and events [8]. These items can be easily transferred to similar ontologies or modified to suit specific needs, thus ensuring reusability of shared knowledge and contributing to interoperability. As a separate metric of assessing Linked Open Data quality reusability factor is taken into account in [23]. For instance, in [5], SPITFIRE model supplements functional SENSEI model by searching for already existing concepts for reuse and datasets to establish linkages.

C. High Level Processing

Large scale of resource heterogeneity and distribution in IoT entails a need for notable computational power. Full potential of IoT will be leveraged by transforming low-level data into high-level data seamlessly and in a comprehensive way [6]. Low-level data are single observation and measurement units captured by sensors or other devices. Usually IoT data consumers, either humans or machines, are not interested in unit measurements but rather in high level contextual meaning of data. Such high-level abstractions should be machine-interpretable for computers and human-understandable for humans. High-level knowledge collection, transfer, data extraction and modification in a specific domain for certain applications are demanding the largest part of processing resources. Currently there are only some solutions that address this problem. For instance, the annotation tool in Sense2Web creates an opportunity for users to annotate sensor data such as resource, entity, and service descriptions [2]. Processing and analysing semantic data sets for gaining information and enabling advanced communication with IoT devices directly depend on effective querying and interaction processing between message brokers [6]. Mature working solutions can already work with large-scale semantic descriptions but a changing environment of IoT demands for even higher processing abilities. Current mechanisms for querying data in semantic technologies are based on SPARQL. Engines tailored for a sensor streaming environment of IoT include Streaming SPARQL, SPARQL Stream and others. All these systems add an extra window for operational RDF stream processing [2]. Experimental work by Qin et al. [24] shows a good result on disseminating Linked Data streams to construct effective user queries. Qin et al. promise that further experimentations will include more sophisticated user queries such as JOIN. Research shows a promising opportunity to support high levels of data consumption needs in the IoT. In [11], Kiljander et al. propose another solution of engine that enables efficient processing of SPARQL subscriptions. One more option is to introduce distributed computation in order to reduce communication overload, termed in network-processing. The idea originates from wireless sensor networks that are set up to

Applied Computer Systems

_____ *2017/21*

perform local measurements [1]. Darbari et al. in [15] develop a granular computing concept that can be used to dynamically cluster information about urban traffic using OWL semantic language in IoT.

Transaction handling is another open issue related to data processing in the IoT [2]. When data is being updated across thousands of devices within a network with high frequency and different update policies, it becomes difficult to define what a transaction is.

Immense data amount generation, although not directly falling under processing categorisation, is also a predictable result of emerging field of IoT [4]. Large data amounts call for intelligent data storage solutions for analytics, monitoring, and actuation. An expected result of smart analytics is automated decision making by means of machine-learning methods based on evolutionary algorithms, neural networks and similar solutions, all of which demand data on-the-go from various storage sources.

D. Data Quality

High level of data quality provides accurate and timely information to manage IoT related services and ensures accountability. It allows measuring service effectiveness and prioritising resource usage. IoT presents an architecture where data is the key factor in both the perception layer and the application layer (see Fig. 2). Data in this case are an actual value perceived by sensors or others devices and passed along for processing; therefore, it is essential to apply the best data quality practices immediately in first steps in data processing in IoT.

Semantic Web development has produced a high volume of data being published on the web as Linked Open Data [23]. Some semantic description models such as the W3C SSN ontology offer attributes to describe qualitative aspects of data [6]. Analysis of such data quality reveals that these data often are not consistent [23]. As stated, data extracted from partly structured or non-structured sources such as DBpedia might contain inaccuracies and misrepresentations. For human interpretation, inaccurate data might be sufficient. For machine-learning or knowledge-dependent application, however, such data might be unacceptable.

While many methodologies and frameworks exist for assessing data quality and deriving appropriate conclusions and suggestions for improvement, the Web of Data contains innovative aspects that are not always covered by existing mechanisms. As data volume increases, inconsistency and redundancy become troubling issues. Inconsistent data detection for distributed networks becomes even more challenging [2]. Quality of observations and measurements can vary over time due to environmental changes, device faults, or system downtimes [6]. Possible methods to detect errors and analyse data in accordance with the defined quality requirements include anomaly detection and filtering, as well as reliable semantic description filtering. Mechanisms of verifying quality and reliability have to support selected semantic approaches.

Several studies focus on quality aspects of semantic technology and IoT data combined, implying that this issue is of current interest of academic community. Zaveri et al. [23], [27] formalise terminology related to data quality and provide a framework for comparing quality assessment approaches and tools. Regarding quality metric definition and validation of tools of quality assessment, it has been concluded that Flemming's data quality assessment tool has covered most of defined quality dimensions [27]. Recent study by Fürber [16] offers another data quality principle classification with a deeper focus on ontologies and links between them. It is proposed to segregate data aspects by means of quality of data source, raw data, semantic conversion, and linking. Survey on context-aware computing by Perera et al. [8] names such quality criteria as quality of physical sensors, context data, and delivery process.

E. Data Confidentiality and Privacy

By means of big data analytics and web mining techniques, behaviour of user can be analysed to improve structure and content of offered services. The helping a user to feel more comfortable and enabling intuitive browsing uncover methods and approaches of utilising user data for malicious means, thus threatening user privacy and confidentiality. Determining threatening behaviour contexts is a ceaseless challenge. Large networks consisting of autonomous heterogeneous devices as in IoT transfer large amounts of valuable information. These kinds of systems will not only own and record information about users but will also produce sensitive data that are rich in context [20]. Health-care is an exquisite field of interest for IoT applications, but the lack of appropriate methods for ensuring privacy and sensitive information has hindered the adoption of IoT developments in health-care [2]. Moreover, in IoT wireless communication will play an important role. Adoption of wireless medium as a standardised environment for data exchange may pose new threats in terms of privacy violation. Hence, such risks should be mitigated within ongoing research to ensure stable development of IoT paradigm. Measures providing the network and device resilience and means of interception of attacks, data authentication, access control and client privacy have to be established [22]. IoT attributes such as globality, technicity, ubiquity, and scalability have to be covered by the proposed solutions. Security and privacy need to be protected at all IoT architecture layers [8].

Certain studies primarily focus on privacy challenges of IoT and imply introduction of legal frameworks, as the study by Weber [22]. Resembling challenges of standardisation, the proposed framework would call for consensus between involved parties to become fully operational as well. Regarding privacy the following open challenges can be identified [2]: defining general model of privacy in IoT, development of innovative enforcement methods, and development of methods that are able to balance the need for confidential data for personalisation needs without overexposing the sensitive information. Mobile reasoning can be used to preserve privacy [10]. Survey papers on IoT vastly

cover the topic of privacy in future Internet networks, as in [1] and [2]. Some on-going initiatives of IoT systems consider privacy and confidentiality concerns within the development of pervasive context aware IoT applications.

F. Data Interpretation and Synthesis

Providing smart objects with interpretation and analytics methods to process and evaluate events in their surroundings is important for building the concept of intelligent machines [6]. Semantic descriptions serve the purpose of transforming large amount of observed and perceived data created by users and machines into high-level concepts that are meaningful for establishing automated decision making processes. However, the non-human perception contributes to existing pool of challenges in IoT. Similar to problems faced by the artificial intelligence research community, in IoT the challenges are data integration and amalgamation from different sources, rules of data aggregation, defining borders and thresholds, as well as describing events, actors and objects. Solutions are needed to integrate data from various environments, register and act upon anomalies (including errors) and patterns for further fusion of new knowledge based on learnt rules [6]. Interpretation challenge is a vaguely noticed topic. Some studies indicate this challenge, but propose no solutions. From Qin et al. [2], it can be derived that interpretation problems can be eliminated by strong semantic annotations. Similar ideas are also expressed in [27]. Tsai et al. [20] predict a future framework of IoT that includes an interpretation layer as a separate mechanism.

V. CONCLUSION

The goal of the survey has been to identify commonalities and open issues regarding the application of semantic technologies in the IoT domain. The paper has provided an analysis of 23 sources by organising the revealed challenges in 6 groups, namely: scalability and flexibility, standardisation and reusability, high level processing, data quality, data confidentiality and privacy, and data interpretation and synthesis.

Merging semantic solutions with different aspects of IoT proves to be a difficult task. The emerging issue of scalability is detected in most of the reviewed publications as it reflects the ever-expanding nature of IoT network. Semantic service-oriented approach is a proper solution that promotes increased scalability for IoT and requires automated dynamic composition mechanisms. Standardisation and reusability can be achieved by introducing common frameworks and reference models, but cooperation of vendor and regulatory parties is an essential topic here. Real-time data stream exchange and handling of the immense data amount are not supported by existing semantic engines, thus creating a need for more efficient mechanisms to be implemented for the IoT paradigm. Approximately half of articles within the survey scope refer to data quality as an issue for semantics in IoT. Defining quality metrics is a key task for future IoT solutions. Data confidentiality principles are also often mentioned. By increasing the number of sensors and devices in IoT networks, the amount of sensitive data transferred rises as well. Some

works offer a deep study of privacy issues and introduce a legal framework as a reference model. Data synthesis and interpretation are an underlying topic for developing IoT networks.

The present survey shows that many aspects of semantics application in the IoT domain are not yet resolved and call for continued research and development. Most of the identified issues are related to the dynamic and pervasive nature of IoT. The challenges are outlined here to encourage further research in the application of semantic technologies in the IoT domain.

REFERENCES

[1] D. Miorandi, S. Sicari, F. De Pellegrini and I. Chlamtac, "Internet of Things: Vision, Applications and Research Challenges," Ad Hoc Networks. vol. 10, no. 7, pp. 1497–1516, 2012. https://doi.org/10.1016/j.adhoc.2012.02.016

[2] Y. Qin, Q.Z. Sheng, N.J.G. Falkner, S. Dustdar, H. Wang and A.V. Vasilakos, "When Things Matter: A Survey on Data-Centric Internet of Things," J. Netw. Comput. Appl., vol. 64, pp. 137–153, 2016. https://doi.org/10.1016/j.jnca.2015.12.016

[3] C. Perera, A. Zaslavsky, P. Christen and D. Georgakopoulos, *CA4IOT: Context Awareness for Internet of Things. In: Proc. of 2012 IEEE Int. Conf. Green Comput. Commun. GreenCom 2012, Conf. Internet Things, iThings 2012 Conf. Cyber, Phys. Soc. Comput. CPSCom 2012,* pp. 775–782, 2012.

[4] J. Gubbi, R. Buyya and S. Marusic, "Internet of Things (IoT): A Vision, Architectural Elements, and Future Directions," Future Gener. Comput. Syst. vol. 29, no. 7, pp. 1645–1660, 2013. https://doi.org/10.1016/j.future.2013.01.010

[5] D. Pfisterer, D. Pfisterer, K. Romer, D. Bimschas, O. Kleine, R. Mietz, C. Truong, H. Hasemann, A. Kröller, M. Pagel, M. Hauswirth, M. Karnstedt, M. Leggieri, A. Passant, R. Richardson, "SPITFIRE: Toward a Semantic Web of Things," Commun. Mag. vol. 49, no. 11, pp. 40–48, IEEE, 2011.

[6] P. Barnaghi, W. Wang, C. Henson and K. Taylor, "Semantics for the Internet of Things," Int. J. Semant. Web Inf. Syst. vol. 8, no. 1, pp. 1–21 2012. https://doi.org/10.4018/jswis.2012010101   D. Boyd and K. Crawford, "Six Provocations for Big Data," Computer and Information Science, vol. 123, 2011.

[8] C. Perera, A. Zaslavsky, P. Christen and D. Georgakopoulos, "Context Aware Computing for the Internet of Things: A Survey," IEEE Commun. Surv. Tutorials, vol. 16, no. 1, pp. 414–454, 2014. https://doi.org/10.1109/SURV.2013.042313.00197

[9] K. Kotis and A. Katasonov, "Semantic Interoperability on the Internet of Things," Int. J. Distrib. Syst. Technol., Igi-Global, vol. 4, no. 3, pp. 47–69, 2013.

[10] A. Maarala, X. Su and J. Riekki, "Semantic Reasoning for Context-ware Internet of Things Applications," IEEE Internet of Things Journal. vol. PP, no. 99, pp. 1–12, 2016.

[11] J. Kiljander, A. D'elia and F. Morandi, "Semantic Interoperability Architecture for Pervasive Computing and Internet of Things," IEEE Access, vol. 2, pp. 856–873, 2014. https://doi.org/10.1109/ACCESS.2014.2347992

[12] W. Wang, S. De, R. Toenjes, E. Reetz and K. Moessner, *A Comprehensive Ontology for Knowledge Representation in the Internet of Things. In: Proc. 11th IEEE Int. Conf. Trust. Secur. Priv. Comput. Commun. Trust. – 11th IEEE Int. Conf. Ubiquitous Comput. Commun. IUCC-2012,* pp. 1793–1798, IEEE, 2012. https://doi.org/10.1109/trustcom.2012.20

[13] B. Christophe, "Federated Semantic Nodes to Scale Search Process in the IoT," Commun. Mob. Comput. vol. 1, no. 1, p. 8, 2012. https://doi.org/10.1186/2192-1121-1-8

[14] Z. Song, A.A. Cárdenas and R. Masuoka, "Semantic Middleware for the Internet of Things," Internet Things, IoT 2010, 2010.

[15] M. Darbari, D. Yagyasen and A. Tiwari, "Intelligent Traffic Monitoring Using Internet of Things (IoT) with Semantic Web. In: Emerging ICT for Bridging the Future - Proceedings of the 49th Annual Convention of the Computer Society of India (CSI)," vol. 1, pp. 455–462, Springer International Publishing 2015. https://doi.org/10.1007/978-3-319-13728-5_51

[16] C. Fürber, Data Quality in the Semantic Web. Data Quality Management with Semantic Technologies, 1st ed., pp. 69–77, Gabler Verlag, 2016. https://doi.org/10.1007/978-3-658-12225-6_5

[17] J. Euzenat, "Research Challenges and Perspectives of the Semantic Web," IEEE Intell. Syst. vol. 17, no. 5, pp. 86–88, 2002. https://doi.org/10.1109/MIS.2002.1039838

[18] J.L. Martinez-Rodriguez, I. Lopez-Arevalo and A.B. Rios-Alvarado, *A Classification of Challenges in the Semantic Web Based on the General Architecture. In: Proc. - Int. Work. Database Expert Syst. Appl. DEXA*, pp. 197–201, 2016.

[19] N. Noy, "Semantic Integration: A Survey of Ontology-Based Approaches," SIGMOD Rec., vol. 33, no. 4, pp. 65–70, 2004. https://doi.org/10.1145/1041410.1041421

[20] C.-W. Tsai, C.-F. Lai and A.V. Vasilakos, "Future Internet of Things: Open Issues and Challenges," Wirel. Networks, vol. 20, no. 8, pp. 2201–2217, 2014. https://doi.org/10.1007/s11276-014-0731-0

[21] *A. Gyrard, S.K. Datta, C. Bonnet and K. Boudaoud, Standardizing Generic Cross-Domain Applications in Internet of Things. 2014 IEEE Globecom Work. GC Workshops 2014,* pp. 589–594, IEEE, 2014.

[22] R.H. Weber, "Internet of Things – New Security and Privacy Challenges," Comput. Law Secur. Rev. vol. 26, no. 1, pp. 23–30, 2010. https://doi.org/10.1016/j.clsr.2009.11.008

[23] A. Zaveri, A. Rula, A. Maurino, R. Pietrobon, J. Lehmann and S. Auer, "Quality Assessment for Linked Open Data: A Survey. Semant," Web J. by IOS Press., vol. 1, pp. 1–31, 2014.

[24] Y. Qin, Q.Z. Sheng and E. Curry, "Matching Over Linked Data Streams in the Internet of Things," IEEE Internet Comput. vol. 19, no. 3, pp. 21–27, 2015. https://doi.org/10.1109/MIC.2015.29

[25] T. Berners-Lee and M. Fischetti, Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor. 1st ed., Harper San Francisco, 1999.

[26] S. De, P. Barnaghi, M. Bauer and S. Meissner, Service Modelling for the Internet of Things. 2011 Federated Conference on Computer Science and Information Systems FedCSIS, Szczecin, pp. 949–955, 2011.

[27] A. Zaveri, A. Rula, A. Maurino, R. Pietrobon, J. Lehmann and S. Auer, "Quality Assessment Methodologies for Linked Open Data: A Systematic Literature Review and Conceptual Framework. Semant. Web – Interoperability, Usability, Appl. vol. 1, p. 33, 2012.

Marina Harlamova, born 1993 in Latvia, is a second-year Master student majoring in Business Informatics, Riga Technical University (2016–2018). The obtained degrees: Bachelor of Engineering Sciences, Riga Technical University (2011–2015).
At present, she is a Business Analyst at Tele2 Shared Service Centre. Fields of interests include computer science, CRM & billing systems, solution design, problem solving and requirements analysis.
E-mail: harlamova.mar@gmail.com

Mārīte Kirikova, *Dr. sc. ing*., is a Professor in Information Systems Design at the Department of Artificial Intelligence and Systems Engineering, Faculty of Computer Science and Information Technology, Riga Technical University, Latvia. She has more than 150 publications on the topics of requirements engineering, business process modelling, knowledge management, and systems development. She has done fieldwork at Stockholm University and Royal Institute of Technology, Sweden, Copenhagen University, Denmark, and Boise University, USA. In her research, currently she focuses on continuous information systems engineering in the context of agile and viable systems approaches.
E-mail: marite.kirikova@gmail.com

Kurt Sandkuhl, born 1963 in Germany, is a full Professor of Business Information Systems at the University of Rostock (Germany) and has an adjunct position of Professor of Information Engineering at the School of Engineering, Jönköping University. He received a diploma (*Dipl.-Inform.*) and a degree of Doctor of Computer Science (*Dr.-Ing.*) from Berlin University of Technology in 1988 and 1994, respectively. He received the title of Associate Professor (postdoctoral lecturing qualification) at Linköping University, Institute of Technology, in 2005. Sandkuhl's current research interests include the fields of information logistics, enterprise modelling, ontology engineering, and model-based software engineering. He has published three books in the field of electronic publishing and more than 150 papers in information logistics, enterprise knowledge management, CSCW, information services, and software architectures.
E-mail: kurt.sandkuhl@uni-rostock.de